# Channel Fragmentation in Dynamic Spectrum Access Systems - a Theoretical Study

Ed Coffman†, Philippe Robert*, Florian Simatos*, Shuzo Tarumi†, Gil Zussman†

†Electrical Engineering
Columbia University

*INRIA Paris-Rocquencourt

{egc,shuzo,gil}@ee.columbia.edu, {philippe.robert,florian.simatos}@inria.fr

## ABSTRACT

Dynamic Spectrum Access systems exploit temporarily available spectrum ('white spaces') and can spread transmissions over a number of non-contiguous sub-channels. Such methods are highly beneficial in terms of spectrum utilization. However, excessive fragmentation degrades performance and hence off-sets the benefits. Thus, there is a need to study these processes so as to determine how to ensure acceptable levels of fragmentation. Hence, we present experimental and analytical results derived from a mathematical model. We model a system operating at capacity serving requests for bandwidth by assigning a collection of gaps (sub-channels) with no limitations on the fragment size. Our main theoretical result shows that even if fragments can be arbitrarily small, the system does not degrade with time. Namely, the average total number of fragments remains bounded. Within the very difficult class of dynamic fragmentation models (including models of storage fragmentation), this result appears to be the first of its kind. Extensive experimental results describe behavior, at times unexpected, of fragmentation under different algorithms. Our model also applies to dynamic linked-list storage allocation, and provides a novel analysis in that domain. We prove that, interestingly, the 50% rule of the classical (non-fragmented) allocation model carries over to our model. Overall, the paper provides insights into the potential behavior of practical fragmentation algorithms.

**Categories and Subject Descriptors:** C.2.1 [Computer-Communication Networks]: Network Architecture and Design — *Wireless communication*; G.3 [Probability and Statistics]: Markov Processes

**General Terms:** Theory, Performance

**Keywords:** Dynamic Spectrum Access, Fragmentation, Ergodicity of Markov chains, Cognitive Radio

## 1. INTRODUCTION

This paper focuses on dynamic resource allocation algorithms in Dynamic Spectrum Access Networks (also known as Cognitive Radio Networks). A Cognitive Radio is a concept that was first defined by Mitola [21, 22] as a radio that can adapt its transmitter parameters to the environment in which it operates. Technically, it is based on the concept of Software Defined Radio [4] that can alter parameters such as frequency band, transmission power, and modulation scheme through changes in software. According to the Federal Communications Commission (FCC), a large portion of the assigned spectrum is used only sporadically [10]. Because of their adaptability and capability to utilize the wireless spectrum opportunistically, algorithms for Dynamic Spectrum Access are key enablers to efficient use of the spectrum. Hence, the potential of Cognitive Radio Networks has been recently identified by various policy, research, and standardization organizations [8, 11, 16].

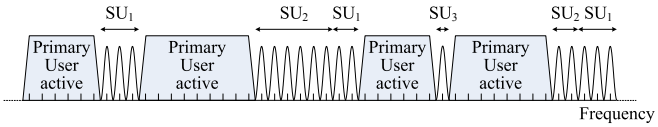



**Figure 1: Non-contiguous OFDM with Primary and Secondary Users (SUs), where the Secondary Users use non-contiguous channels that do not overlap with the Primary Users' channels.**

Under the basic model of Cognitive Radio Networks [2], Secondary Users (SUs) can use *white spaces* (also known as *spectrum holes*) that are not used by the Primary Users but must avoid interfering with active Primary Users (e.g., Figure 1). For example, the Primary and Secondary Users can be viewed as TV broadcasters and cellular operators using available TV bands [16]. Under this model, one assumes that when a transmission of a Primary User takes place, it occupies a predefined band. An SU may identify spectrum holes (not used by Primary or other Secondary Users) and can allocate its bandwidth among a number of subchannels, occupying a number of holes (not necessarily contiguous). This can be realized, for example, by employing a variant of Orthogonal Frequency-Division Multiplexing (OFDM) that is capable of deactivating sub-carriers which have the potential to cause interference to other users [14, 20, 23–25] (such a non-contiguous OFDM scheme is shown in Figure 1).

The use of non-contiguous bandwidth blocks results in non-trivial behavior even for very simple scenarios. As an example, due to the dynamic use of the available spectrum holes and the arrivals and departures of SU bandwidth requests, SUs may need to transmit in a set of smaller and


*SIGMETRICS'10*, June 14–18, 2010, New York, New York, USA.

smaller holes. This will lead to a highly fragmented spectrum whose maintenance may require complex algorithmic solutions. Although the practical (physical layer) aspects of OFDM-based Dynamic Spectrum Access have been extensively studied recently, the use of fragmented (i.e., non-contiguous) spectrum introduces several new problems [17, 24,26] that significantly differ from classical Medium Access Control (MAC) and fragmentation problems.

In this paper, we study the most *basic theoretical model* in which the spectrum is shared by SUs only. Those users have to transmit and receive data, and accordingly need some bandwidth for given amounts of time. Hence, bandwidth requests of SUs are characterized by a desired total bandwidth and the duration of a time interval over which it is needed. The data transmission can take place over a non-contiguous channel (i.e., a number of subchannels). Once a transmission terminates, some fragments (subchannels) are vacated, and therefore, gaps (spectrum holes/white spaces) develop randomly in both size and position.

When allocating a channel to a new SU request, it is being fragmented (in the frequency domain) into available gaps until the full requested bandwidth is provided. This process repeats itself, until the next request fails to fit into the available fragments. Figure 2 demonstrates the process of a transmission termination followed by the fragmentation of waiting bandwidth requests (more details about this example can be found in Section 2). We note that from a practical point of view, such a system can be viewed as a simplistic version of an OFDM-based access-point/base-station/spectrum-broker that tracks the available spectrum holes and allocates non-contiguous bandwidth blocks to SUs.

The main goal of the paper is to investigate the phenomena of fragmentation induced by spectrum allocation algorithms. For this purpose, we will ignore the particularities of techniques such as OFDM and make a couple of assumptions (for a complete system description, see Section 2): (i) the system operates at capacity and there is always a waiting bandwidth request; and (ii) the fragment size is not bounded from below. Making the first assumption allows us to study the effect of fragmentation in the worst case. Clearly, if there are idle periods, when there are no waiting requests, only departures occur and the fragmentation level of the system decreases during these periods. Similarly, the latter assumption allows us to study the system performance when artificial lower bounds on the fragment size are not imposed. This differs from OFDM-based systems in which a subcarrier has a given minimal bandwidth.

## Main Results

At first glance, it seems that given assumption (ii) above, as time passes, the fragments used by bandwidth requests might well become progressively narrower and that the number of fragments grows without bound. Clearly, such an operation model cannot be supported by a realistic system (e.g., an OFDM-based system). However, our analytical and experimental results show that *this is not the case*.

In particular, under our model, a portion of the spectrum (represented by the interval $[0, 1]$) is available for allocation to requests whose sizes are uniform on $(0, \alpha]$ $(0 < \alpha \leq 1)$. Once allocated, requests are satisfied (leave the system) after an exponential time. A stochastic model of channel fragmentation is investigated which includes the asymptotic behavior of the number $G(t)$ of *gaps*, $F(t)$ of *fragments*, and $R(t)$ of requests being served (also referred to as allocated *channels*). A Markovian description of this system is clearly more complicated than the vector $(G(t), F(t), R(t))$, since for each of the $R(t)$ channels, the location and size of each of its fragments should be included in such a description.

Let $(t_n)$ be the sequence of request departure times. Our main result shows that the total number of gaps and fragments at these epochs has bounded exponential moments,

$$\sup_{n\geq 1} \mathbb{E}\left(e^{\eta(F(t_n)+G(t_n))}\right) < +\infty,$$

for some $\eta > 0$. It follows that the sum $F(t_n) + G(t_n)$ is strongly concentrated near the origin, indicating that, *with high probability, low fragmentation of the spectrum holds under high traffic intensity*. In our analysis, several basic ingredients are used: relations between the number of fragments and gaps in the spectrum; a drift relation of Lyapunov type for the total number of gaps, fragments, and requests (Proposition 1); a general inequality for Markov chains (Theorem 4); and a stability result of [18].

In addition to the analytical results, and in order to gain insight into the performance of the system, we present the results of an extensive simulation study. We show that although for a given maximum request size $(\alpha)$, the number of fragments has a finite expected value, there is a linear relationship between $1/\alpha$ and the expected number of fragments into which a request is divided. This indicates that when the requests are very small, they are also fragmented into a relatively large number of fragments. From a practical point of view, this implies that *imposing a lower bound on the fragment size is important* in order to reduce the complexity resulting from maintaining a channel heavily fragmented.

We show that different bandwidth allocation algorithms have significantly different performance in terms of the fragmentation occurring in the system. In particular, an algorithm (referred to as LFS) that allocates gaps in decreasing order of their sizes reduces the fragmentation by almost an order of magnitude compared to other algorithms. Interestingly, we also show that the number of fragments is distributed according to a Normal law with a relatively low mean value. Finally, we observe an unexpected reappearance of the *50% rule* [19], according to which, on average the number of gaps (unoccupied fragments) is half the number of requests being served (this rule differs from the original 50% rule by Knuth [19] which is briefly discussed below).

## Related Work

The areas of Dynamic Spectrum Access and Cognitive Radio Networks have been extensively studied (for a comprehensive review of previous work see [1, 2]). In particular, problems stemming from fragmented spectrum have been considered in [17,24,26]. In addition, several previous works focused on the technical aspects of channel fragmentation [13, 20, 23, 25]. To the best of our knowledge, despite these recent efforts, the fragmentation problem considered in this paper has not been studied before.

Yet, the spectrum allocation problems described here can be characterized in the context of classical dynamic storage allocation problems of computers, see Knuth [19]. In that context, the term "spectrum" refers to the storage unit, "bandwidth" refers to storage, and "channel" refers to a region in storage containing a file. In the original storage model, fragmentation refers only to the gaps of unoccupied

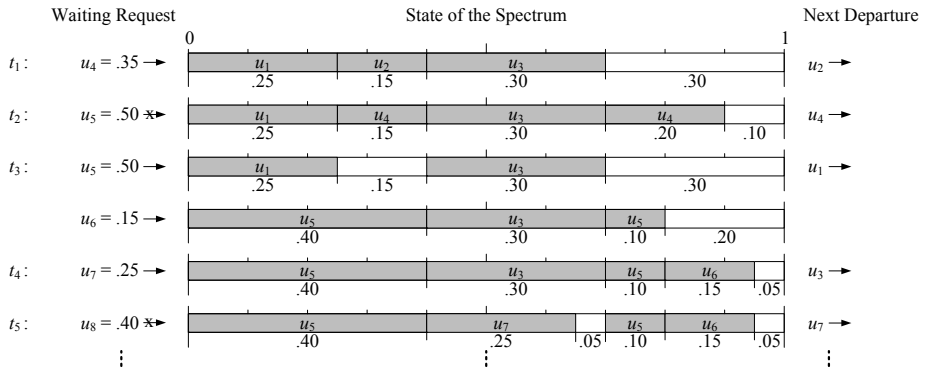



**Figure 2: An example of the admission and departure processes and the resulting spectrum fragmentation.**

storage interspersed with intervals of occupied storage: files are not fragmented. The 50% rule was derived for this model and asserted that the number of gaps was approximately half the number of files when exact fits of files into gaps were rare. Our notion of fragmentation applies also to the files themselves, so our discovery that the 50% rule continued to apply was a surprising one at first. Note that, in terms of the storage application, our model corresponds well to *linked-list* allocation of files – channels allocated to requests are sets of linked, disjoint segments of storage allocated to specific file fragments. Our results provide a novel analysis for such systems, and have implications for the *garbage-collection/defragmentation* process in linked-list systems.

We note that studies of dynamic storage allocation have been around for some 40 years, and widely recognized as posing very challenging problems to both combinatorial and stochastic modeling and analysis. In particular, results of the type found in this paper, rigorous within stochastic models, seem to be quite new.

For a system without fragmentation, the early results of Kipnis and Robert [18] concern a stochastic analysis of the number of requests being served in the system. Fragmentation is not an issue in their case since request allocations can be moved when needed in order to put available space all together in one block. A major result in [18] applying to our model asserts the existence and uniqueness of an invariant measure for the number of requests. Explicit formulas for our system are hard to come by, but those in [18] for the maximal departure rate in special cases have provided useful checks on our experiments.

## Organization of the Paper

The remainder of this paper is organized as follows. Section 2 describes the system, allocation algorithm, and the mathematical variables describing fragmentation. Section 3 presents simulation results that bring out the behavior of the system and the effects of fragmentation. Section 4 introduces some relations between the variables describing the fragmentation and a *50% limit law* for the relation between the number of active channels and the number of gaps in the spectrum at departure times. Section 5 contains our main mathematical result, Theorem 2, which shows that the average value of the total number of fragments and gaps is bounded. Furthermore, in some cases, the existence of an equilibrium distribution is established in Theorem 3. Section 6 discusses algorithmic issues, such as changes in performance resulting from alternative algorithms for sequencing through the list of gaps when constructing a channel for a newly admitted request. Section 7 presents experiments suggesting that Normal approximations for the total number of fragments and gaps hold. Section 8 discusses the results and future research directions. Due to space constraints, some of the details are omitted and can be found in [5].

# 2. THE MODEL

## System Model

As mentioned above, we consider requests for bandwidth queued up waiting to be served, each of them identified with the amount of bandwidth required and a time telling how long it wants a channel with that total bandwidth. Channels are allocated to bandwidth requests on a FCFS (first-come-first-served) basis, subject to available bandwidth; the channels must remain fixed while active; they depart after varying delays, so *gaps* alternating with sequences of subchannels (channel *fragments*) develop over time.

For convenience, we normalize the spectrum to the interval $[0, 1]$, so all bandwidth requests are numbers in $[0, 1]$. There is a queue of waiting requests that never empties. Assuming that the spectrum initially has no channels in use, the allocation process begins by allocating consecutive channels i.e., consecutive subintervals of [0,1], to requests whose sizes are $u_1, u_2, \ldots$ until a request whose size $u_i$, $i > 1$, is reached which exceeds available bandwidth, i.e.,

$$u_1 + \cdots + u_{i-1} \leq 1 < u_1 + \cdots + u_i$$

All $i - 1$ of the channels now begin their independent delays. Subsequent state transitions take place at departure epochs when the delays of currently allocated channels expire. At such epochs, all fragments of the departing channel are released. Suppose all requests up to $u_j$ have been allocated channels and a request $u_i$, $i \leq j$, departs, releasing its allocated channel. Then $u_{j+1}, u_{j+2}, \ldots$ are allocated their requested bandwidths until, once again, a request is encountered that asks for more bandwidth than is available. All channels then begin or continue their delays until the next departure. A standard rule for setting up a channel is to scan the spectrum, from one end to the other, with gaps of available bandwidth allocated to fragments of the new requested bandwidth until enough has been allocated to satisfy the entire request. The last gap used in satisfying a request is normally only partially used; the partial alloca-

tion in the last gap is left-justified in that gap. We refer to this scanning rule as *Linear Scan*.

An example is shown in Figure 2. Allocations for the first 8 user requests whose sizes are $u_1, \cdots, u_8$ are shown, assuming that, at time 0, the spectrum is not in use. The requests with sizes $u_1$, $u_2$, and $u_3$ are the first to be allocated channels; $u_4$ must wait for a departure, since the first 4 request sizes sum to more than 1. The variables $t_i$ give the sequence of departure times of allocated requests. We see that the first occurrence of fragmentation takes place at the departure of $u_2$ and the subsequent admission of $u_4$; an initial fragment of $u_4$ is placed in the gap left by $u_2$ and a final fragment is placed after $u_3$. Note also that, even after $u_2$ and $u_4$ have departed, there is still not enough bandwidth for $u_5$. After the additional departure of $u_1$, both $u_5$ and $u_6$, but not $u_7$, can be allocated bandwidth.

In Section 6, we shall also evaluate two alternatives to the *Linear Scan* rule: a *Circular Scan*, by which each linear scan starts where the previous one left off, and a *Largest-First Scan* intended to further reduce fragmentation. Although these algorithms make different scans of the gap sequence, they are all alike in their treatment of the last gap occupied: *the last fragment is left justified in the last gap.* This is a key assumption, and it is very likely to hold in practice.

The allocation process described in this section contrasts with the classical model of dynamic storage allocation, in which each bandwidth request must be accommodated by a single, sufficiently large gap of available spectrum. However, the process does correspond closely to the linked-list model of dynamic storage allocation, in which an available-space list is maintained and files can be fragmented in accordance with this list. This paper also yields a novel stochastic analysis of such systems. Note that the assumption that there is always another request in queue models a system operating at capacity, where the departure rate, which is equal to the admission rate, is often called maximum throughput.

## Notation, State Space, and Probability Model

We denote by $U$ the size of the request waiting to be allocated bandwidth, and by $U_i$ for $i \geq 1$ the size of the $i$th request behind it. Except in the proof of Theorem 3 in Section 5, we omit the dependency in time of these variables for the sake of notation. As mentioned in Section 2, we denote by $R(t)$ the number of channels (requests that are allocated bandwidth) at time $t$. $F(t)$ and $G(t)$ denote the number of fragments and gaps at time $t$. A Markovian description of this system is clearly more complicated than the vector $(G(t), F(t), R(t))$. Hence, we now define the state space of the fragmentation process. At a given state, we denote by $r$ the number of channels (requests to which bandwidth has been allocated) and by $s_i$ $(1 \leq i \leq r)$ the amount of bandwidth allocated to these requests. A state of the spectrum $[0, 1]$ carries the information given by a sequence in which gaps alternate with sets of contiguous fragments:

DEFINITION 1. *A state $x$ in the state space $\mathcal{S}$ of the fragmentation process is denoted by*

$$x = (L_1, \ldots, L_r; u)$$

*where $u$ is the size of (amount of bandwidth required by) the request waiting at the head of the queue and $L_i$, is the list of open subintervals of $[0, 1]$ occupied by the fragments of the $i$-th channel. For $x$ to be admissible, the open intervals in $\cup_i L_i$ must be mutually disjoint, and, since the size of $u$ exceeds the bandwidth available, $u > 1 - \sum_i s_i$ has to hold.*

Note that for a specific state, the size of the requests in the queue are denoted by $u$ and $u_j, j \geq 1$ and the size of the channels (requests that are already allocated bandwidth) are denoted by $s_i, 1 \leq i \leq r$.[1]

Let $(X(t))$ be the process living in the state space $\mathcal{S}$. In the *probability model* used in this paper, the bandwidth requests $(U_i)$ are independent random variables which are uniformly distributed on $(0, \alpha], 0 < \alpha \leq 1$, and request residence times are i.i.d. mean-1 exponentials.[2] Under this probability model, $(X(t))$ is a Markov process on $\mathcal{S}$.

# 3. EXPERIMENTAL RESULTS

In this section we describe an experimental study of the model described above. This serves two related roles. First, it brings out characteristics of the fragmentation process that need to be borne in mind in implementations, particularly where these characteristics show conditions (e.g., parameter settings) that must be avoided, if a system with fragmentation is to operate efficiently. The second role is that of experimental mathematics, in which results indicate where behavior might well be formalized and rigorously proved as a contribution to mathematical foundations. In the latter role, this section leads up to the next two sections, which formalize the stability of the fragmentation process.

The experiments were conducted with a discrete-event simulator written in C. In general, the tool is capable of simulating stochastic request arrival/departure processes. For this paper, however, the arrival process was effectively inoperative, since the interest here is behavior while the system is operating at capacity (i.e., at maximum throughput). Admissions are made whenever the waiting request fits into available spectrum, and once made the waiting request is immediately replaced by another with independent samples from the required-bandwidth and residence-time distributions. The admission process continues in this way, effectively simulating a queue that never empties, until no further admissions can be made and one or more departures need to occur. The excellent accuracy of the tool was established in tests against exact queueing results, and exact results from [18]. Due to space constraints, the verification details are omitted and can be found in [5].

Recall that bandwidth requests are uniformly distributed on $(0, \alpha], 0 < \alpha \leq 1$. Hence, the principal parameter of the experimental model will be $\alpha$. The simulations of stationary behavior were most demanding, of course, for small $\alpha$. For every $\alpha$ value, $0.01 \leq \alpha \leq 1$, 20 million departure events were simulated starting in an empty state, with data collected for the last 10 million events. For $0.001 \leq \alpha \leq 0.01$, 100 million departure events were processed and data collection was performed during the last 50 million events.

The results for the average number of channels, the average number of gaps, and the average number of fragments per channel are shown in Figure 3. The curves are nearly linear in $1/\alpha$ (the errors in the linear fits are within the thickness of the printed lines). In particular, the asymptotic

[1]For simplicity of the presentation, we did not use the notation $s_i$ in the example provided in Figure 2.

[2]The assumption that request sizes are uniformly distributed is for convenience. The results hold for more general distributions.

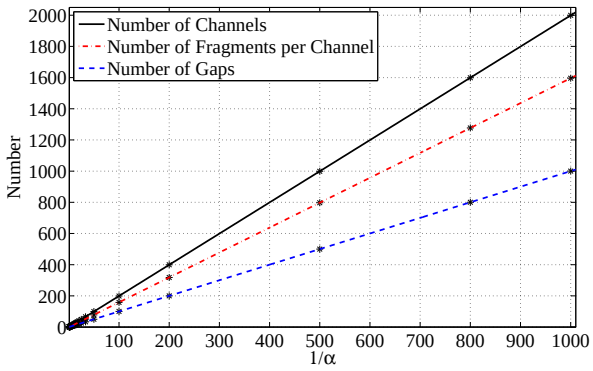



**Figure 3: Average numbers of channels, of gaps, and of fragments per channel.**

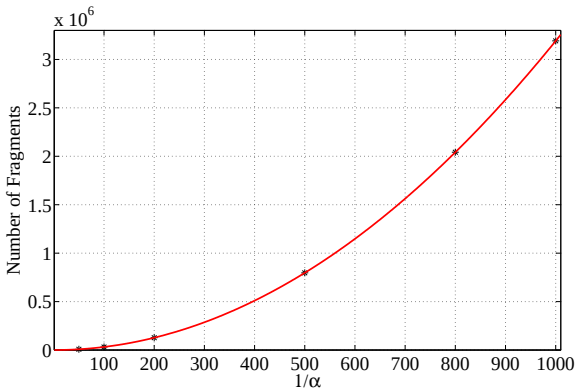
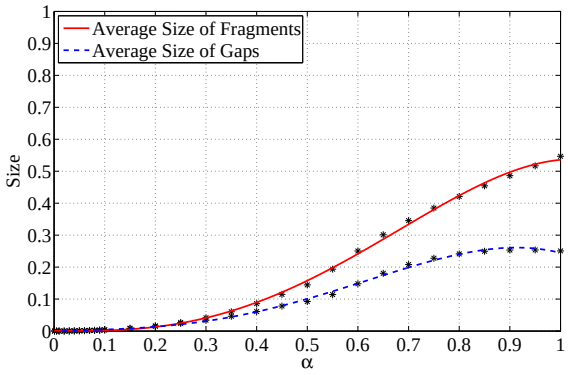
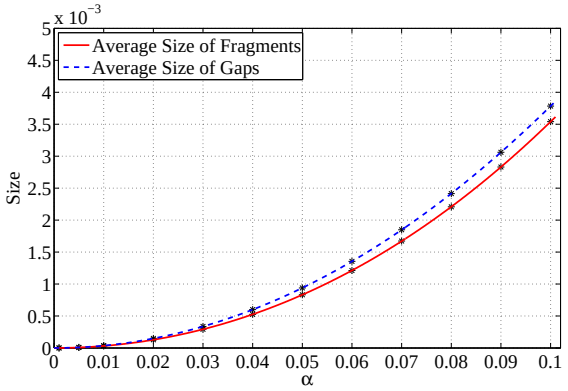



**Figure 4: Average total number of fragments vs. $1/\alpha$: a quadratic fit.**




(a) $0 \leq \alpha \leq 1$




(b) $0 \leq \alpha \leq 0.1$

**Figure 5: Average sizes of fragments and gaps.**

average number of channels in the spectrum is $2/\alpha$. When requests are large relative to the spectrum (i.e., for $\alpha > 1/3$), the behavior is not given by functions quite so simple. As such cases are of less practical interest, we omit the relevant data due to space constraints.

The asymptotic linear growth of the average number of channels as a function of channel size is obvious, but the linearity of the other two measures is not so obvious. A closer look shows that the average number of gaps is almost exactly one half the average number of channels for even relatively small $1/\alpha$. This is an unexpected version of Knuth's 50% rule for dynamic storage allocation. We return to this behavior in the next section, where we prove a 50% limit law. The linear growth of the average number of fragments per channel may also be unexpected at first glance: the fragmentation of channels *increases* as the average channel size *decreases*. This linear growth implies the quadratic growth of the average total number of fragments plotted in Figure 4 (the accuracy of the fit is as before: the error is within the thickness of the printed lines).

The analysis in the later sections will focus largely on tracking fragment *types* defined as follows: a fragment is of type-$i$, if it is adjacent to $0, 1,$ or 2 fragments. In [5] we provide additional simulation results and show that for small $\alpha$ more than 90% of the fragments are type-2 fragments. In addition, clearly, the number of type-0 and type-1 fragments is a function of the number of gaps. These observations and the results illustrated in Figures 3 and 4 indicate that, even for relatively small $1/\alpha$, the average total number of type-0 and type-1 fragments grows linearly in $1/\alpha$, but the average number of type-2 fragments grows quadratically.

Figure 5 compares the average gap and fragment sizes. As might be expected, for relatively small $\alpha$, they are close to each other. The relation holds even for moderately large $\alpha$, although for $\alpha$ rather close to 1, the difference amounts to about a factor of 2. With this property and the 50% rule suggested by Figure 3, the linear growth in the number of fragments per channel (shown in Figure 3) is easily explained for moderately small $\alpha$ in the following way.

As mentioned above, for moderately small $\alpha$, the number of channels is approximately $2/\alpha$ (i.e., the spectrum size divided by the average request size). Due to the 50% rule, the number of gaps is roughly $1/\alpha$. At any time, the total size of the gaps is at most $\alpha$, since there is a request waiting for departures whose requested bandwidth exceeds that total size of gaps. Therefore, at most $\alpha$ available bandwidth is spread among $1/\alpha$ gaps, giving an average gap size on the order of $\alpha^2$. The average fragment size is at most (and indeed very close to) the average gap size. The fragments must occupy at least $1-\alpha$ of the spectrum (since, as mentioned above, at most a fraction $\alpha$ of the spectrum is devoted to gaps). Thus, the number of fragments must be on the order of $1/\alpha^2$, and so the average number of fragments per channel must be on

the order of (in particular, linear in) $1/\alpha$. As $\alpha \to 0$, the asymptotics of these estimates become more precise.

# 4. NUMBERS OF FRAGMENTS AND GAPS

As mentioned in Section 2, under our probability model, $(X(t))$ (the process living in the state space $\mathcal{S}$) is a Markov process on $\mathcal{S}$. In this section, we obtain analytical results regarding the number of fragments and gaps under this process. We begin by formally defining the fragment types which were discussed in Section 3.

DEFINITION 2. *For $i = 0,1,$ or 2, a fragment is said to be of type $i$ if it touches exactly $i$ other fragments. $N_i(t)$ denotes the number of type $i$ fragments at time $t$, so that $F(t) = N_0(t) + N_1(t) + N_2(t)$ is the total number of fragments.*

Let $\sigma(t)$ denotes the sum of the number of fragments and gaps,

$$\sigma(t) = F(t) + G(t). \tag{1}$$

The number $G(t)$ of gaps and the numbers of fragment types are related as follows.

LEMMA 1. *With probability* 1,

$$G(t) = N_0(t) + \frac{1}{2}N_1(t) + I(t) \tag{2}$$

*For any $t \geq 0$, where $I(t) = 1$, if there is a gap starting at the origin, and 0 otherwise.*

PROOF. Each gap, except for boundary gaps starting at 0 or ending at 1, separates two fragments. Two gaps surround a type-0 fragment not touching the origin, only one touches a type-1 fragment not touching the origin, and none touch a type-2 fragment, so the gaps strictly inside $(0,1)$ are double-counted in $2N_0(t) + N_1(t)$. Gaps at the boundaries are counted only once in this expression, so if there are gaps touching each boundary, 2 must be added to $2N_0(t) + N_1(t)$ to produce a double count of all gaps. Then $N_0(t) + N_1(t)/2 + 1$ counts the gaps as called for by the lemma.

With probability 1, a gap always touches the boundary at 1, so the only case left to consider is the absence of a gap touching the origin. In this case, there is a type-0 or type-1 fragment touching the origin, and so a nonexistent gap has been counted in $2N_0(t) + N_1(t)$. This over-count cancels the under-count of the gap touching 1, and so no correction term is needed, i.e., $N_0(t) + N_1(t)/2$ counts all gaps as stated in the lemma. □

DEFINITION 3. *Let $(t_k)$ denote the sequence of departure times, let $D_i(t_k)$ denote the number of type $i$ fragments in the channel leaving at time $t_k$, and let $A(t_k)$ denote the number of requests admitted to the spectrum at time $t_k$. Finally, define the* drift *in the total number of fragments and gaps:*

$$\Delta\sigma(t_k) \stackrel{def}{=} \sigma(t_k) - \sigma(t_{k-1}). \tag{3}$$

The following lemma is the basis of the stability analysis of $\sigma(t)$ in Section 5, and the 50% rule proved later in this section.

LEMMA 2. *With probability* 1, *the departure at $t_k$ creates the following change in the total number of fragments and gaps:*

$$\Delta\sigma(t_k) = A(t_k) - 2D_0(t_k) - D_1(t_k) + J(t_k), \quad k \geq 1 \tag{4}$$

*with $t_0 = 0$, and $J(t_k) = 1$, if a fragment starting at the origin is in the departing channel, and 0 otherwise.*

PROOF. With probability 1, each new channel allocation covers completely every gap it is allocated, except for the last one, which is only partially covered. Thus, with probability 1, each new channel allocation changes gaps to fragments, except for the last gap which is changed to a fragment plus a gap; this adds one to $\sigma(t_{k-1})$ for each admission, which accounts for the total of $A(t_k)$ in (4).

Two fragments of the same channel can not be contiguous, so it is correct to add up the changes created by departing fragments, with each being treated separately. Suppose first that there is no fragment $(0,b)$ against the origin. Then for every type-0 fragment in the departing channel, two gaps and a fragment are replaced by a single gap for a net decrease of two, and for every departing type-1 fragment, a gap and a fragment are replaced by a single gap for a net reduction of one. This gives the reduction of $2D_0(t_k) + D_1(t_k)$ appearing in (4). If there is a fragment $(0,b)$, it must be of type 0 or 1; if it is of type 0, then its departure gives a decrease of one; if it is of type 1, its departure has no effect. Each of these contributions is one less than it would be were the fragment not touching the origin. There can only be one such fragment, so the correction shown in $J(t_k)$ for a fragment $(0,b)$ follows. □

We will denote by $G^-(t_k)$ the total number of gaps just after the $k$-th departure, but before new admissions, if any, are made. Note that if we remove $A(t_k)$ from the right-hand side of (4) and add back the total number of departing fragments at $t_k$, i.e., $D_0(t_k)+D_1(t_k)+D_2(t_k)$, we get the number of gaps available to admissions at the $k$-th departure:

$$G^-(t_k) = G(t_{k-1}) - D_0(t_k) + D_2(t_k) + J(t_k) \tag{5}$$

with $J(t_k) = 1$, if there is a departing fragment $(0,b)$, and 0 otherwise.

Knuth's widely known *50% rule* appears in a very different context than the model here, so it is difficult to anticipate the apparent fact that it also holds for our fragmentation model. However, one can argue a similar result assuming only that the fragmentation process has a stationary distribution. The result is given below as an expected value of a ratio, rather than a ratio of expected values.

THEOREM 1. *Assume that the fragmentation process at departure epochs $t_k$ has the stationary distribution $\pi_\alpha$, then as $\alpha \to 0$,*

$$\mathbb{E}_{\pi_\alpha}\left(\frac{G(t_k)}{R(t_k)}\right) \sim \frac{1}{2}$$

PROOF. In the stationary regime, one has, by Lemma 2,

$$\mathbb{E}_{\pi_\alpha}[\sigma(t_k) - \sigma(t_{k-1})] =$$
$$\mathbb{E}_{\pi_\alpha}[A(t_k) - 2D_0(t_k) - D_1(t_k) + J(t_k)] = 0,$$

and to balance departure and admission rates, we must have $\mathbb{E}_{\pi_\alpha} A(t_k) = 1$. Thus, we can write

$$\mathbb{E}_{\pi_\alpha}[2D_0(t_k) + D_1(t_k) - J(t_k)] = 1. \tag{6}$$

Now for a state $x$ at time $t_k$ having $r$ channels and $N_i(t_k)$ type-$i$ ($i = 0,1,2$) fragments, we have from Lemma 1

$$\mathbb{E}_x[2D_0(t_k) + D_1(t_k) - J(t_k)] =$$
$$(2N_0(t_k) + N_1(t_k))/r - \mathbb{E}_x[J(t_k)] =$$
$$2[G(t_k) - I(t_k)]/r - \mathbb{E}_x[J(t_k)] = 2G(t_k)/r + O(1/r)$$

Thus, dropping the $O(1/r)$ term that tends to 0 with $\alpha$ uniformly in $x$, the expectation in (6) proves the theorem. □

## 5. STABILITY RESULTS

This section establishes that the average total number of fragments and gaps remains bounded and that, for certain distributions of request sizes, ergodicity holds. The analysis leads to the two following results. Recall that $(t_n)$ is the sequence of departure times ($t_0 = 0$) and $\sigma(t) = F(t)+G(t)$, defined in (1), is the total number of fragments and gaps.

THEOREM 2. *There exists some $\eta > 0$ such that, for any initial state $x \in \mathcal{S}$,*

$$\sup_{n\geq 1} \mathbb{E}_x\left(e^{\eta\sigma(t_n)}\right) < +\infty. \tag{7}$$

Clearly, this implies that for any initial state $x \in \mathcal{S}$, the sequence $(\mathbb{E}_x(\sigma(t_n)), n \geq 0)$ is bounded. With an additional assumption on the distribution of the request size, a stronger stability result can be proved.

THEOREM 3. *When $\alpha > 1/2$, the process $(X(t))$ is positive Harris recurrent; in particular, it has a unique stationary distribution.*

A criterion for finite exponential moments using a Lyapunov function is established next. Then, we provide some estimates of the drift of the number of fragments between departures which will show us how to construct a Lyapunov function. After constructing this function, we will be in position to prove the boundedness of exponential moments.

### A Criterion for Finite Exponential Moments

Before stating the main result, some results on Markov chains are needed. In the sequel, $\leq_{st}$ refers to stochastic ordering, i.e., $V \leq_{st} Z$ means that $\mathbb{E}(f(V)) \leq \mathbb{E}(f(Z))$ for any increasing function $f$. For reasons that will become clear in Lemma 5, the following lemma focuses on admissions at 4 consecutive departure times. Recall that $(U_i, i \geq 1)$ are the sizes of the requests waiting to be allocated bandwidth (after the first one $U$), which are assumed to i.i.d.

LEMMA 3. *The random variable $A(t_1) + \cdots + A(t_4)$ is stochastically dominated by a random variable $Z$ such that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$.*

PROOF. It is clear that $A(t_1) \leq Z + 1$ where

$$Z = 1 + \inf\{n \geq 1 : U_1 + \cdots + U_n \geq 1\}.$$

The Markov inequality shows that for any $z \geq 0$,

$$\mathbb{P}(Z \geq z + 1) = \mathbb{P}(U_1 + \cdots + U_z \leq 1) \leq e\left(\mathbb{E}\left(e^{-U_1}\right)\right)^z$$

and so $\mathbb{E}(e^{\eta Z})$ is finite for $\eta > 0$ small enough. From this observation, it is not difficult to extend the result to $A(t_1) + \cdots + A(t_4)$ instead of just $A(t_1)$. □

This lemma shows in particular that

$$\xi \stackrel{\text{def}}{=} \sup_{i\geq 1} \sup_{x\in\mathcal{S}} \mathbb{E}_x(A(t_i)) < +\infty.$$

is well-defined; this constant will be used repeatedly throughout the rest of the analysis. The proof of the following lemma is standard, and therefore, omitted.

LEMMA 4. *Let $Z \geq 0$ be a positive, real-valued random variable such that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$, and define $c = \lambda^{-2}\mathbb{E}(e^{\lambda Z} - 1 - \lambda Z)$. Then for any $0 \leq \varepsilon \leq \lambda$ and any real-valued random variable $V$ such that $V \leq_{st} Z$, we have $\mathbb{E}\left(e^{\varepsilon V}\right) \leq 1 + \varepsilon\mathbb{E}(V) + \varepsilon^2 c$.*

The following result is closely related to result of Hajek [15]. For completeness its proof can be found in [5].

THEOREM 4. *Let $(Y_k)$ be a discrete-time, continuous state-space Markov chain such that for some function $f \geq 0$, there exist $K, \gamma > 0$ such that for any initial state $y$ with $f(y) > K$, $\mathbb{E}_y(f(Y_1) - f(Y_0)) \leq -\gamma$. Assume that there exists a random variable $Z$ such that for any initial state $y$, $Z$ dominates stochastically the random variable $f(Y_1) - f(Y_0)$ under $\mathbb{P}_y$. Assume finally that $\mathbb{E}(e^{\lambda Z}) < +\infty$ for some $\lambda > 0$. Then there exist $\eta > 0$ and $0 \leq C < +\infty$ such that for any initial state $y$,*

$$\sup_{n\geq 1} \mathbb{E}_y\left(e^{\eta f(Y_n)}\right) \leq e^{\eta f(y)} + C.$$

Theorem 4 will be applied to the Markov chain $(X(t_{4n}))$ with a function $f$ of the form $\sigma_\kappa = \sigma + \kappa r$ for some $\kappa > 0$ suitably chosen. $(X(t_{4n}))$ is not the most natural choice at first glance, but it appears to be needed because of the complexity of the state space.

It is clear that $\sigma(t_1) - \sigma(t_0) \leq A(t_1) + 1$, so that

$$\sigma_\kappa(t_4) - \sigma_\kappa(t_0) \leq (\kappa + 1)(A(t_1) + \cdots + A(t_4)) + 4$$

and therefore, by Lemma 3, $\sigma_\kappa(t_4) - \sigma_\kappa(t_0)$ is stochastically dominated by some random variable $Z$ with an exponential moment. Therefore, one has to establish a negative drift relation for $\sigma_\kappa(t_4) - \sigma_\kappa(t_0)$. This is the purpose of the following two subsections.

### Evolution of the Number of Fragments

Recall that $x \in \mathcal{S}$, the initial state of the system, has $r$ active channels, and define the total available gap size $h = 1 - (s_1 + \cdots + s_r)$. Time 0 referring to the initial state $x$ will usually be omitted; e.g., $\sigma(0), F(0), G(0), \ldots$ will be simplified to $\sigma, F, G, \ldots$. Recall that $\Delta\sigma(t_n)$ is defined in (3) as $\sigma(t_n) - \sigma(t_{n-1})$.

LEMMA 5. *Fix $0 < \varepsilon < 1$ and $0 < \eta < 1/2$, and let $x \in \mathcal{S}$ be an initial state such that $\sigma = G + F \geq 2K + 1$ for some fixed $K \geq 0$.*

*Then $F = N_0 + N_1 + N_2 \geq K$, and*

1) *If $r = 1$, then $\mathbb{E}_x(\Delta\sigma(t_1)) \leq \xi - K$.*

2) *If $r > 1$ and $N_0 + N_1 \geq \varepsilon K$, then*

$$\mathbb{E}_x(\Delta\sigma(t_1)) \leq \xi + 1 - \frac{\varepsilon K}{r}.$$

*Assume in the remaining cases that $r > 1$, define $K' = K\big((1-\varepsilon)/r - \varepsilon\big)^+$, and let $i^* \in \{1, \ldots, r\}$ index a channel $L_{i^*}$ in $x$ with the most type-2 fragments.*

3) *If $N_0 + N_1 \leq \varepsilon K$ and $u > h + s_{i^*}$, then*

$$\mathbb{E}_x(\Delta\sigma(t_2)) \leq \xi + 2 - \frac{K'}{r(r-1)}. \tag{8}$$

4) *If $N_0 + N_1 \leq \varepsilon K$, $u < h + s_{i^*}$ and $h + s_{i^*} < \eta\alpha$, then*

$$\mathbb{E}_x(\Delta\sigma(t_3)) \leq \xi + 2 - \frac{(1-\eta)K'}{r^2(r-1)}. \tag{9}$$

5) *If $N_0 + N_1 \leq \varepsilon K$, $u < h + s_{i^*}$ and $\eta\alpha < h + s_{i^*}$, then there exists a $\gamma(\eta) > 0$ such that*

$$\mathbb{E}_x(\Delta\sigma(t_4)) \leq \xi + 2 - \frac{\gamma(\eta)K'}{r^5}. \tag{10}$$

*It follows that there exists a $\overline{\xi} > 0$ and a function $\psi(r) > 0$ such that for any $x$ with $\sigma \geq 2K+1$,*

$$\mathbb{E}_x(\sigma(t_4) - \sigma) \leq \overline{\xi} - K\psi(r). \tag{11}$$

PROOF. As is readily verified, $G \leq F+1$, so $2K+1 \leq \sigma = F+G \leq 2F+1$, and hence $F \geq K$ as claimed. In what follows, we use repeatedly the two following simple facts:

$$\mathbb{E}_x\left(D_0(t_1) + D_1(t_1)\right) = (N_0+N_1)/r, \tag{12}$$

and by Lemma 1,

$$G \geq K \Rightarrow N_0 + N_1 \geq K-1. \tag{13}$$

— *First case:* $r = 1$. Then, right after the only channel initially present leaves, there is no channel allocated bandwidth, and therefore, $\sigma(t_1) = A(t_1)$. Note that $r = 1$ is only possible when $\alpha > 1/2$, and in this case the possibility for a channel to be alone is crucial in the proof of the Harris recurrence stated in Theorem 3.
— *Second case:* $r > 1$, $N_0 + N_1 \geq \varepsilon K$. Then the inequality follows from (4):

$$\begin{aligned}\mathbb{E}_x(\Delta\sigma(t_1)) &\leq \xi + 1 - \mathbb{E}_x(D_0(t_1) + D_1(t_1)) \\ &= \xi + 1 - \frac{N_0+N_1}{r} \leq \xi + 1 - \frac{\varepsilon K}{r}.\end{aligned}$$

In the 3 remaining cases, let $N_j^*$ denote the number of type-$j$ fragments in any channel $i^*$ which has the most type-2 fragments. If $N_0 + N_1 \leq \varepsilon K$, then since $F \geq K$, necessarily $N_2 \geq (1-\varepsilon)K$ and $N_2^* \geq (1-\varepsilon)K/r$. Define the event $D^* = \{\text{channel } L_{i^*} \text{ leaves at } t_1\}$ and recall that $G^-$ denotes the number of gaps right after $L_{i^*}$ leaves but before new admissions, if any, are made. It follows from (5) that $G^- \geq K'$ in the event $D^*$, since

$$G^- = G - N_0^* + N_1^* + J(t_1) \geq (-\varepsilon K + (1-\varepsilon)K/r)^+ = K'.$$

The remaining analysis tacitly assumes that $r > 1$, that $N_0 + N_1 \leq \varepsilon K$ and that the channel $L_{i^*}$ leaves at $t_1$.
— *Third case:* $u > h + s_{i^*}$. Then $A(t_1) = 0$, since when $L_{i^*}$ leaves it does not provide enough additional bandwidth for $U$. In particular, $R(t_1) = r-1$ and $G(t_1) = G^- \geq K'$, and so

$$\mathbb{E}_x(\Delta\sigma(t_2)) \leq \xi + 1 - \mathbb{E}_x(D_0(t_2) + D_1(t_2); D^*).$$

The strong Markov property makes it possible to lower-bound this last term.

$$\begin{aligned}&\mathbb{E}_x(D_0(t_2)+D_1(t_2); D^*) = \mathbb{E}_x(\mathbb{E}_{X(t_1)}(D_0(t_1)+D_1(t_1)); D^*) \\ &= \mathbb{E}_x\left(\frac{(N_1+N_2)(t_1)}{R(t_1)}; D^*\right) \geq \frac{K'-1}{r-1}\mathbb{P}_x(D^*) = \frac{K'-1}{r(r-1)}\end{aligned}$$

and therefore, $\mathbb{E}_x(\Delta\sigma(t_2)) \leq \xi + 2 - K'/(r(r-1))$.
— *Fourth case:* $u < h + s_{i^*} < \eta\alpha$. In this case $U$ is admitted at $t_1$. Thus it makes sense to define the event

$$E_4 = D^* \cap \{U \text{ leaves at } t_2 \text{ and } U_1 > \eta\alpha\}.$$

Then as before

$$\mathbb{E}_x(\Delta\sigma(t_3)) \leq \xi + 1 - \mathbb{E}_x(D_0(t_3) + D_1(t_3); E_4).$$

In the event $E_4$, $U$ is admitted at $t_1$ and leaves at $t_2$, while $U_1$ stays blocked at $t_1$ and $t_2$, so that $G(t_2) = G^- \geq K'$ and $R(t_2) = r-1$. Hence as in the second case,

$$\mathbb{E}_x(D_0(t_3) + D_1(t_3); E_4) \geq \frac{K'-1}{r-1}\mathbb{P}_x(E_4) \geq \frac{(1-\eta)K'}{r^2(r-1)} - 1$$

since $\mathbb{P}_x(E_4) = (1-\eta)/r^2$. Thus (9) holds.
— *Fifth case:* $u < h + s_{i^*}$ *and* $\eta\alpha < h + s_{i^*}$. Again, $U$ is admitted at $t_1$. Letting $U_i$ denote the sizes of the requests behind $U$, define the event

$$B = \{U_i < \eta\alpha,\ i = 1, \ldots, \tau \text{ and } U_{\tau+1} > 2\eta\alpha\}$$

with $\tau = \inf\{n \geq 0 : U_1 + \cdots + U_n > h + s_{i^*} - \eta\alpha\}$ and $E_5' = D^* \cap B \cap \{U \text{ leaves at } t_2\}$. It is readily verified that $1 \leq \tau < +\infty$ almost surely. Moreover, one has in $E_5'$

$$0 < h^* \stackrel{\text{def}}{=} h + s_{i^*} - (U_1 + \cdots + U_\tau) < \eta\alpha < U_{\tau+1}.$$

This means that at $t_2$, exactly $\tau$ new requests $U_1, \ldots, U_\tau$ have been admitted, and $U_{\tau+1}$ is blocked. Moreover, for any $i \in \{1, \ldots, \tau\}$, one has $h^* + U_i < 2\eta\alpha < U_{\tau+1}$, so that if one of the $\tau$ channels allocated to the $(U_i)$ leaves, $U_{\tau+1}$ remains blocked.

When $L_{i^*}$ left, there were $G^- \geq K'$ gaps; in the remainder of the analysis, we call an *initial gap* a gap present right after $L_{i^*}$ left. After $L_{i^*}$ left, $U$ and $A(t_1) - 1$ new requests were admitted, and then $U$ left and $A(t_2)$ new requests were admitted at $t_2$. Thus, at $t_2$, each initial gap is in either of two states: either it is completely filled, or it is still a gap, i.e., it has not been filled completely. Let $k$ be the number of initial gaps completely filled at $t_2$, and let $k' = G^- - k$: then $k + k' = G^- \geq K'$. In each initial gap completely covered at $t_2$, there is at least one type-2 fragment of one of the $\tau$ new channels. Therefore, $N_{1,2} + N_{2,2} + \cdots + N_{\tau,2} \geq k$ with $N_{i,2}$ the number of type-2 fragments of the channel corresponding to $U$. In particular there is a channel $L_{j^*}$, $j^* \in \{1, \ldots, \tau\}$ with at least the average $k/\tau$ of type-2 fragments: $N_{j^*,2} \geq k/\tau$. Define finally the event $E_5 = E_5' \cap \{L_{j^*} \text{ leaves at } t_3\}$. Since $h^* + U_{j^*} < U_{\tau+1}$, then $U_{\tau+1}$ remains blocked at $t_3$ when $E_5$ occurs, and therefore (note that when $j^*$ leaves, some gaps may merge, but not two initial gaps),

$$G(t_3) \geq N_{j^*,2} + k' \geq k/\tau + k' \geq (k+k')/\tau \geq K'/\tau.$$

Now we proceed as before, to obtain

$$\mathbb{E}_x(\Delta\sigma(t_4)) \leq \xi + 1 - \mathbb{E}_x(D_0(t_4) + D_1(t_4); E_5)$$

and, using the Markov property at time $t_3$,

$$\begin{aligned}&\mathbb{E}_x(D_0(t_4)+D_1(t_4); E_5) = \mathbb{E}_x(\mathbb{E}_{X(t_3)}(D_0(t_1)+D_1(t_1)); E_5) \\ &= \mathbb{E}_x\left(\frac{(N_0+N_2)(t_3)}{R(t_3)}; E_5\right) \geq \mathbb{E}_x\left(\frac{(K'/\tau-1)^+}{r+\tau-2}; E_5\right)\end{aligned}$$

since $R(t_3) = r + \tau - 2$ in $E_5$. The same kind of reasoning as before then leads to

$$\mathbb{E}_x\left(\frac{(K'/\tau-1)^+}{r+\tau-2}; E_5\right) \geq \frac{K'}{r^5} f(\eta, h + s_{i^*} - \alpha\eta) - 1$$

with the function $f(\eta, \cdot)$ defined for $y > 0$ by

$$f(\eta, y) = \mathbb{E}\left((1+\tau(y))^{-5}; B(\eta, y)\right)$$

with $\tau(y) = \inf\{n \geq 1 : U_1 + \cdots + U_n \geq y\}$ and

$$B(\eta, y) = \{U_i < \eta\alpha, i = 1, \ldots, \tau(y) \text{ and } U_{\tau(y)+1} > 2\eta\alpha\}.$$

It is not difficult to show that $\gamma(\eta) = \inf_{0<y<1} f(\eta, y) > 0$ which then gives the result.

It remains to prove (11). One only needs to assemble the various bounds, taking into account that $\mathbb{E}_x(\Delta\sigma(t_i)) \leq \xi + 1$ for any $x \in \mathcal{S}$ and $i \geq 0$, to arrive at 4 separate bounds on $\mathbb{E}_x(\sigma(t_4) - \sigma)$. For example, using the former bound for

the first two terms and the last term of $\mathbb{E}_x(\sigma(t_4)-\sigma) = \sum_{1\le i\le 4}\mathbb{E}_x\Delta\sigma(t_i)$ and then the bound in (9) for the third term, we get that one of the 4 bounds, which applies when $x$ satisfies the inequalities of the fourth case, is

$$\mathbb{E}_x(\sigma(t_4)-\sigma) \le 4\xi + 5 - \frac{(1-\eta)K'}{r^5}$$

Computing the minimum over these bounds with $\eta = 1/4$ and $\varepsilon = 1/r^2$, one obtains (11) after setting $\overline{\xi} = 4\xi + 5$ and

$$\psi(r) = \frac{\varphi(r)}{r^6} \times ((1-\eta)\wedge\gamma)$$

with $\varphi(r) = 1 - 2r^{-2}$. This concludes the proof. □

We turn now to the case where $r$ is large. In this case, the negative drift comes from the fact that, except perhaps at $t_1$, with high probability there is no admission at a departure, since the channel that leaves is small with high probability. However, we see in (4) that this is not enough for $\sigma$ to decay, for one would need at least one type-0 or type-1 fragment to leave as well, and it can be the case that most fragments are of type-2. The second term of the Lyapunov function allows us to get around this problem. Since the variation $\Delta R(t_k)$ in the number of channels at a departure is exactly equal to $A(t_k) - 1$, one readily gets that

$$\sigma_\kappa(t_4) - \sigma_\kappa = (\sigma(t_4)-\sigma) + \kappa(A(t_1) + \cdots + A(t_4) - 4).$$

In particular, if $x \in \mathcal{S}$ is such that $\sigma \ge 2K+1$ and $r \le K_r$, then (from now on, we assume without loss of generality that the function $\psi$ given by (11) in Lemma 5 is decreasing)

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \le \overline{\xi} - K\psi(K_r) + 4\kappa(\xi - 1)$$

whereas if $r \ge K_r$,

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \le 4(\xi+1) + \kappa\mathbb{E}_x(r(t_4) - r)$$

and so we see that we only need to control $\mathbb{E}_x(r(t_4)-r)$ for $r$ large.

LEMMA 6. *There exist $K_r, \gamma_r > 0$ such that if $x \in \mathcal{S}$ is such that $r \ge K_r$, then $\mathbb{E}_x(r(t_4) - r) \le -\gamma_r$.*

PROOF. Since the technical difficulty of the proof of this inequality is similar to that of the above proof, we need only give a sketch of it. From $s_1 + \cdots s_r = 1 - h \le 1$ one gets $\#\{i : s_i \ge \gamma\} \le 1/\gamma$, and therefore, $\mathbb{P}_x(s_{i_1} \ge 1/\sqrt{r}) \le 1/\sqrt{r}$ with $L_{i_1}$, $i_1 \in \{1,\ldots,r\}$, the channel that leaves at $t_1$. Thus, when $r$ is large, with high probability a small channel leaves.

If $h-u$ is away from $\alpha$, then the event $\{u_1 > h - u + s_{i_1} + \cdots + s_{i_4}\}$ (with $i_k$ defined similarly) has high probability, and in this event $A(t_1) + \cdots + A(t_4) \le 1$. If in contrast $h-u$ is large, then with high probability $U_1$ is admitted and with high probability $h - u - U_1$ is away from $\alpha$; hence we can do the same again, and get that, with high probability, $A(t_1) + \cdots + A(t_4) \le 2$. The lemma is proved. □

## Construction of a Lyapunov Function

For $\kappa > 0$, one defines, for an initial state $x \in \mathcal{S}$ of the system, $\sigma_\kappa \stackrel{\text{def}}{=} \sigma + \kappa\, r$, where $r$ is the number of channels allocated in $x$ and $\sigma = F + G$ is the sum of the number of fragments and gaps in $x$.

PROPOSITION 1. (Lyapunov function inequality) *There exist $\kappa$ and $K > 0$ such that if $x \in \mathcal{S}$ is such that $\sigma_\kappa \ge K$, then $\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \le -1$.*

PROOF. Let $K_r$ and $\gamma_r$ be as in Lemma 6, and take $\kappa$ and $K$ as follows:

$$\kappa = \frac{4\xi+5}{\gamma_r} \quad \text{and} \quad K = \frac{8\xi + 2\overline{\xi} + 2}{\psi(K_r)} + \kappa K_r + 1.$$

Assume that $\sigma_\kappa \ge K$. If $r \ge K_r$, then

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \le 4(\xi+1) - \kappa\gamma_r = -1.$$

Otherwise, $r \le K_r$, and since $\sigma_\kappa \ge K$, this necessarily gives $\sigma \ge K - \kappa K_r = 2\hat{K} + 1$ with $\hat{K} = (K - \kappa K_r - 1)/2$. Thus,

$$\mathbb{E}_x(\sigma_\kappa(t_4) - \sigma_\kappa) \le \overline{\xi} - \hat{K}\psi(K_r) + 4\xi = -1$$

and the proposition follows. □

## Proof of Theorem 2

Theorem 4 and Lemma 3 applied to the Markov chain $(X(t_{4n}), n \ge 0)$, and the function $\sigma_\kappa$ show that for some $\eta > 0$ and some constant $0 \le C < +\infty$,

$$\sup_{n\ge 0}\mathbb{E}_x\left(e^{\eta\sigma_\kappa(t_{4n})}\right) \le e^{\eta\sigma_\kappa} + C.$$

Then the Markov property gives for any $i \ge 0$

$$\sup_{n\ge 0}\mathbb{E}_x\left(e^{\eta\sigma_\kappa(t_{4n+i})}\right) \le \mathbb{E}_x\left(e^{\eta\sigma_\kappa(t_i)}\right) + C < +\infty$$

from which (7) follows readily.

## Proof of Theorem 3

In the following discussion, no conceptual argument is missing, only some formalism needed to handle the continuous state space $\mathcal{S}$. These details are routine and left to the interested reader. In the analysis below, requests are said to be *big* if their size exceeds $1/2$.

We argue that $(X(t))$ visits infinitely often a state in which there are no fragmented channels, and such that all size distributions remain the same at all visits. This is enough to show Harris recurrence; see for instance Asmussen [3]. For this purpose, it is convenient to pick a simple regeneration set $E \subset \mathcal{S}$ in which (i) the spectrum is being used by a big request, alone and with an unfragmented channel of the form $(0, b)$, and (ii) the request $U$ waiting at the head of the queue is also big. Each of these has the conditional request-size distribution given that its size is larger than $1/2$, i.e., the uniform distribution on $(1/2, \alpha)$, see [18].

To verify that $E$ is visited infinitely often, consider the process $(R(t), U(t))$ with $R(t)$ the number of requests allocated a channel at time $t$ and $U(t)$ the size of the request at the head of the queue; the process $(R(t), U(t))$ is simply the process $(X(t))$ when the data on fragmentation is ignored. This process is positive Harris recurrent, as shown in Kipnis and Robert [18]. In particular it visits infinitely often states with $R(t) = 1$ and $U(t) > 1/2$; one can add $U_1(t) > 1/2$ as well (i.e., the first request in line behind the head of the queue request is also big), since this happens with a geometric probability. Then when the only channel leaves, the process $(X(t))$ enters $E$, since there is exactly one channel, it is big, it is necessarily unfragmented and of the form $(0, b)$, and a big request is waiting at the head of the queue. Moreover, this argument shows that the time between visits to $E$ is integrable, which in turn establishes positive Harris recurrence. This completes the proof.

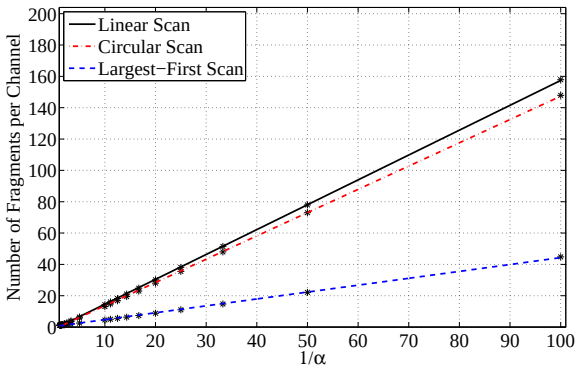



**Figure 6: Average number of fragments per channel.**

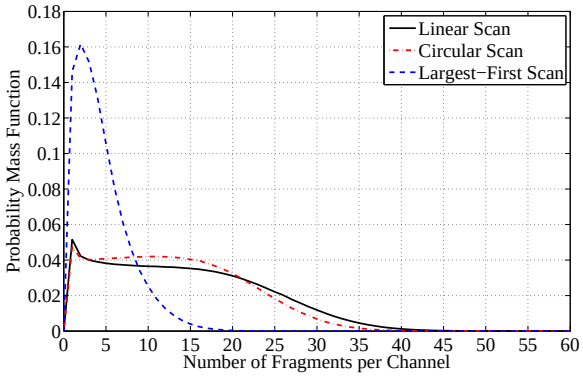



**Figure 7: Distribution of the number of fragments per channel for $\alpha = 0.1$.**

## 6. ALGORITHMS

Although the focus so far has been on measures of fragmentation as a function of $\alpha$, algorithmic issues are also of obvious interest. For example, more uniform patterns of gaps might be an advantage. The *Linear Scan* (LS), discussed in the previous sections, tends to push the gaps towards the end of the spectrum, particularly when the spectrum is viewed at random times in steady state. Interestingly, our experiments have shown that, for all $\alpha < 1/3$, the starting position of the first gap in the spectrum remains very close to 0.64.

To uniformize gap locations, an alternative gap scan resembles the *Circular Scan* (CS) sequences of dynamic storage allocation [19]. In our case, CS uses a circular gap list, in which the successor to the last gap in [0,1] is the first gap in [0,1]. The scan is still linear, but the starting gap of the scan moves as follows: if the last fragment of a channel is placed in gap $g$, then the residual gap of $g$ is the first gap scanned in constructing the next channel. Clearly, although CS will tend to uniformize gap sizes as a function of position, boundary effects will persist so long as the spectrum itself is not circular, i.e., gaps and fragments are not allowed to overlap the end of the spectrum, a restriction that would likely be dictated in practice.

The average number of fragments per channel is a direct measure of gap-search times, and one that we use here. For values of $\alpha$ expected to be of interest in applications, the effects of CS on gap-search times are only within a few percent relative to LS, as can be seen in Figure 6. The figure also shows the average number of fragments per channel for the *Largest-First Scan* (LFS) algorithm. This algorithm is designed to speed up the process of finding a set of gaps sufficient to create a new channel. It selects available gaps in a decreasing order of their sizes and allocates them to a request, thereby greedily minimizing the number of gaps needed to fulfill a request. The extra mechanism needed for such a search will of course tend to reduce overall performance gains. The results in Figure 6 for LFS show a surprisingly large improvement in the average number of fragments per channel – as can be seen, a reduction by a factor more than 3 is achieved by LFS for even moderately small $\alpha$.

The Probability Mass Functions (pmf's) of the number of fragments per channel are shown in Figure 7. Notice that while all probabilities are small under LS and CS, the largest applies to the case of no fragmentation at all. The much more peaked distribution for LFS has both much smaller mean and variance: The standard deviation under CS and LS is approximately 1.5 to 2.0 times that under LFS. The same limiting behavior called for by the 50% law holds for CS and LFS, which was to be expected, as the arguments supporting the 50% law did not depend on the sequence in which gaps were scanned. But an interesting result of our experiments with CS is that the 50% approximation to expectation of the ratios is within a couple of percent *even when the maximum request size is as much as 0.2 the spectrum size*. This is easily seen in Figure 8. The convergence rate of LFS to 50% is intermediate between LS and CS.

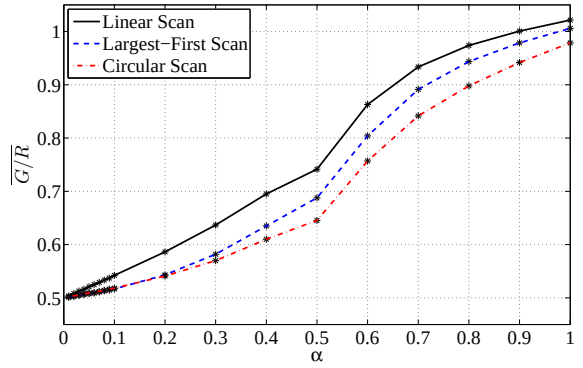
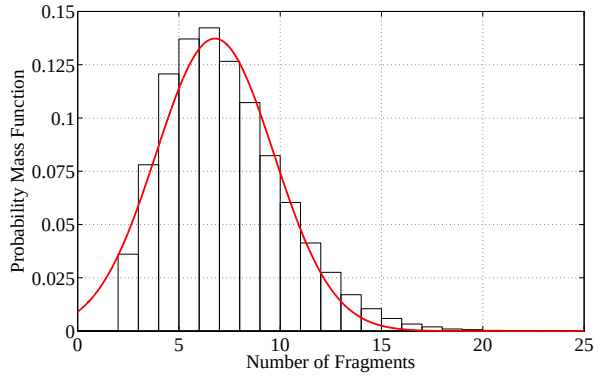



**Figure 8: $\overline{G/R} \to 1/2$ as $\alpha \to 0$ under LS, LFS, and CS.**

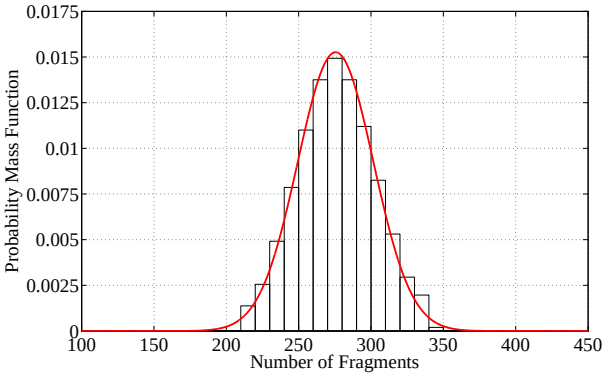
## 7. NORMAL APPROXIMATIONS

After the usual scaling (i.e., first centering then normalizing by the standard deviation), the scaled version of the number of channels, $R$, tends in probability to the standard Normal, $\mathcal{N}(0,1)$, as $M \to \infty$, where $M = \lfloor 1/\alpha \rfloor$ (it is convenient to express the asymptotics in this section in terms of $M$). This result follows easily from the corresponding heavy-traffic limits in [6,7], and the ergodicity of $(R(t))$.

Interestingly, it was discovered in the experiments that a normal limit law also appears to hold for the total number $F$ of fragments as $M \to \infty$ (see for example Figure 9). This is not surprising, since $F$ is the sum over all requests in system of the numbers $F_i$, $1 \le i \le R$, of fragments allocated to




(a) $\alpha = 0.5$




(b) $\alpha = 0.1$

**Figure 9: Distributions of the total number of fragments and corresponding Normal fits.**

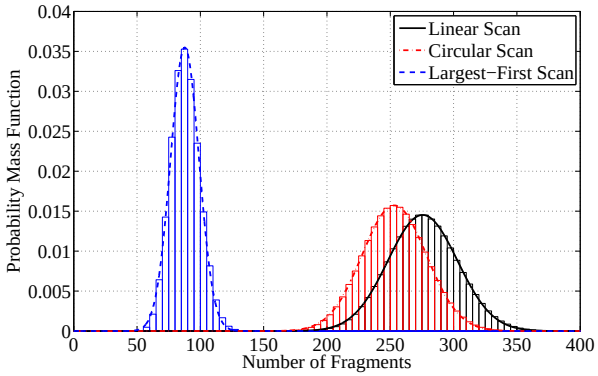



**Figure 10: Distributions of the total number of fragments for $\alpha = 0.1$ and the corresponding Normal fits.**

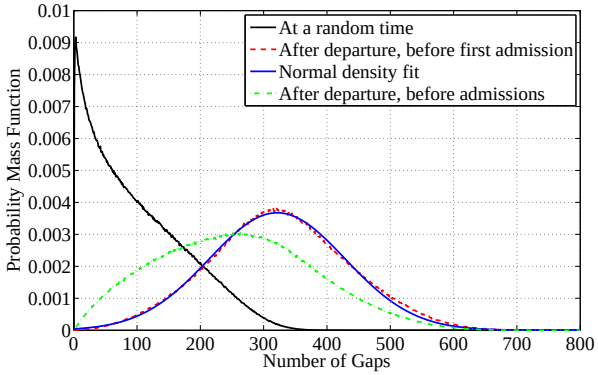



**Figure 11: Distributions of the number of gaps at different epochs for $\alpha = 0.01$.**

requests. The requests have a mutual dependence, but one whose effect can be expected to weaken for large $M$. As illustrated in Figure 6, the mean of $F_i$ is proportional to $M$. Our experiment indicate that the standard deviation of $F_i$ is also approximately proportional to $M$. Hence, if $\xrightarrow{p}$ denotes convergence in probability as $M \to \infty$, then

$$\frac{F - \beta M^2}{\theta M^{3/2}} \xrightarrow{p} \mathcal{N}(0,1) \tag{14}$$

with $\beta \approx 1.5$ and $\theta \approx 0.9$ formalizes the normal-limit law suggested by our experiments.

Note the two departures from the standard Central Limit Theorem set-up. First, consistent with the linear fragmentation observation, individual request fragmentation scales linearly in $M$, so that the total number of fragments scales as $M^2$. Restating the limit law in terms of the sums of random variables $(F_i/M)$ clearly eliminates this discrepancy. Second, the number of channels $(R)$ is random and satisfies the normal limit law discussed above. Thus, the plausibility of (14) requires an appeal to Central Limit Theorems for random sums (e.g., see [12], p. 258, for an appropriate version). Finally, Figure 9 gives some idea of the convergence of the fits to the normal density; as can be seen, fits for $\alpha \leq 0.1$ are indeed close to the simulation data.

We note that the normal approximations shown for the total number of fragments under LS were also found to hold under CS and LFS. This is illustrated in Figure 10.

The distributions of the number of gaps at different epochs are shown in Figure 11. The distribution at a random time shows a decreasing pmf. What appeared to be yet another normal approximation was discovered when looking at the *first-admission* pmf in Figure 11; this is the distribution as seen by the first admission immediately after a departure at just those epochs when there is at least one admission. The third curve is the pmf of the number, $G^-$, of gaps as seen right after departures, but before determining whether or not the head of the queue fits in the total available bandwidth. The fit to the Normal density is also illustrated in the figure. The proof of a similar limit law for Renyi's space-filling problem can be found in [9], but extension of known techniques once again faces the difficult challenges posed by our more difficult fragmentation problem.

## 8. CONCLUSIONS

The results of this paper prepare the ground for further research on several fronts. Before listing a number of the more important ones, we review what we have learned. Our experiments brought out first an unexpected reappearance of a *50% rule* relating the expected numbers of gaps and channels in the limit of small request sizes relative to the spectrum size. In our case, we were able to prove the limit law. The next result described a linear relationship between

the maximum request size, $\alpha$, and the expected number of fragments into which a request was divided at the time of allocation. Interestingly, the smaller $\alpha$ was taken, the greater was the resulting fragmentation of requests.

Our stability results established the beginning of a mathematical foundation of fragmentation processes. Particularly, we showed that for $\alpha > 1/2$, the fragmentation process is Harris recurrent. For general $\alpha$, we proved (with considerable effort) that the total number of fragments is bounded in expected value. We examined alternative algorithms for sequencing through the available gaps and showed that using the LFS algorithm leads to significantly less fragmentation than using the LS or CS algorithms. Finally, we exhibited experimentally a limiting, small-$\alpha$ behavior in which, with appropriate scaling, distributions tend to Normal.

A broad direction for further research extends the parameters of our mathematical model. For instance, while Uniform distributions are generally the assumption of choice in fragmentation models, it would be interesting to see what new effects are created by other distributions of request size, e.g., by varying $a$ in the generalized uniform distributions on $[0, \alpha]$, with densities $x^a/\alpha^{a+1}$. The exponential residence-time assumption is likely to yield simplifications to analysis, but changes in behavior resulting from other distributions are worth investigating. Moreover, instead of a system operating at capacity, in which there is always a request waiting, one could adopt an underlying, fully stochastic model of demand, e.g., a Poisson arrival process, as found in [18].

More realistic, but in all likelihood significantly more difficult models, would relax the independence assumptions. A prime example appropriate for Dynamic Spectrum Access applications would be allowing residence times to depend on fragmentation, the greater the fragmentation of a request, the longer its residence time.

The results regarding the performance of the different algorithms imply that the algorithms' design should also be considered carefully. Some examples of algorithms that come to mind will aim to better fit the fragments into the available gaps. A more challenging objective would be to develop algorithms that take into account spectrum sensing capabilities during the gap allocation process.

Finally, another broad and very important avenue of research that introduces more realistic models discretizes request sizes and the bandwidth allocation process (as is being done while allocating OFDM subcarriers). As in other models of fragmentation, the continuous limit represented in this paper may conceal important effects, or, conversely, it may introduce effects not present in discrete models. We are actively pursuing this avenue of research.

## 9. ACKNOWLEDGMENTS

We would like to thank Charles Bordenave for helpful discussions in relation with Theorem 4. This work was partially supported by NSF grant CNS-0916263 and CIAN NSF ERC under grant EEC-0812072.

## 10. REFERENCES

[1] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury. CRAHNs: Cognitive radio ad hoc networks. *Ad Hoc Networks*, 7(5):810–836, 2009.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Comput. Netw.*, 50(13):2127–2159, Sep. 2006.

[3] S. Asmussen. *Applied Probability and Queues*. Springer, New York, NY, USA, 2nd edition, 1987.

[4] V. G. Bose, A. B. Shah, and M. Ismert. Software radios for wireless networking. In *Proc. IEEE INFOCOM'98*, Apr. 1998.

[5] E. Coffman, P. Robert, F. Simatos, S. Tarumi, and G. Zussman. Channel fragmentation in dynamic spectrum access systems - a theoretical study. Technical Report #2010-03-30, Columbia University, EE, Mar. 2010. http://www.ee.columbia.edu/~zussman/chfrag.pdf.

[6] E. G. Coffman, Jr., A. A. Puhalskii, and M. I. Reiman. Storage limited queues in heavy traffic. *Probab. Eng. and Info. Sci.*, 5(4):499–522, 1991.

[7] E. G. Coffman, Jr. and M. I. Reiman. Diffusion approximations for storage processes in computer systems. In *Proc. ACM SIGMETRICS'83*, 1983.

[8] DARPA XG WG, BBN Technologies. The XG vision RFC version 2.0. 2003.

[9] A. Dvoretsky and H. Robbins. On the 'parking' problem. *Publ. Math. Inst. Hung. Acad. Sci.*, 9:209–226, 1964.

[10] FCC. 03-222, Notice of Proposed Rulemaking, Oct. 2003.

[11] FCC. 08-260, Second Report and Order, ET Docket No. 04-186, Unlicensed Operation in the TV Broadcast Bands, Nov. 2008.

[12] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. John Wiley & Sons, New York, NY, USA, 2nd edition, 1966.

[13] S. Geirhofer, L. Tong, and B. Sadler. Interference-aware OFDMA resource allocation: A predictive approach. In *Proc. IEEE MILCOM'08*, Nov. 2008.

[14] A. Ghasemi and E. Sousa. Spectrum sensing in cognitive radio networks: Requirements, challenges and design trade-offs. *IEEE Commun.*, 46(4):32–39, Apr. 2008.

[15] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Adv. Appl. Probab.*, 14(3):502–525, 1982.

[16] IEEE 802.22, Working Group on Wireless Regional Area Networks ("WRANs"). Documentation available at http://ieee802.org/22/.

[17] J. Jia, Q. Zhang, and X. Shen. HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management. *IEEE J. Sel. Areas Commun.*, 26(1):106–117, Jan. 2008.

[18] C. Kipnis and P. Robert. A dynamic storage process. *Stochastic processes and their applications*, 34(1):155–169, 1990.

[19] D. E. Knuth. *The Art of Computer Programming, Vol. 1 - Fundamental Algorithms*. Addison Wesley Longman Publishing Co., Redwood City, CA, USA, 3rd edition, 1997.

[20] H. Mahmoud, T. Yucek, and H. Arslan. OFDM for cognitive radio: Merits and challenges. *IEEE Wireless Commun.*, 16(2):6–15, Apr. 2009.

[21] J. Mitola III. Cognitive radio for flexible mobile multimedia communications. In *Proc. IEEE MoMuC'99*, Nov. 1999.

[22] J. Mitola III. *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Doctor of Technology, Royal Inst. Technol. (KTH), Stockholm, Sweden, 2000.

[23] J. D. Poston and W. D. Horne. Discontiguous OFDM considerations for dynamic spectrum access in idle TV channels. In *Proc. IEEE DySPAN'05*, Nov. 2005.

[24] A. Shukla, B. Willamson, J. Burns, E. Burbidge, A. Taylor, and D. Robinson. A study for the provision of aggregation of frequency to provide wider bandwidth services. Technical report, QinetiQ, Aug. 2006.

[25] T. A. Weiss and F. K. Jondral. Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency. *IEEE Commun.*, 42(3):S8–14, Mar. 2004.

[26] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu. Allocating dynamic time-spectrum blocks in cognitive radio networks. In *Proc. ACM MobiHoc'07*, 2007.